STIC-ILL

From: Huynh, Phuong N.
Sent: Friday, October 12, 2001 9:12 AM
To: STIC-ILL

NPL only

Good morning!

Please deliver the following:

Gene 1997 Sep 15;197(1-2):277-87

Biochim Biophys Acta 1997 Mar 1;1355(3):303-14

Genomics 1998 Jan 1;47(1):84-91

Thank You,
Neon
Art unit 1644
Mail CM1, 9E12
Office CM1, 9D06
Tel 308-4844

GENOMICS 47, 84–91 (1998)
ARTICLE NO. GE975085

The Mouse Chondroadherin Gene: Characterization and Chromosomal Localization

Christina Landgren,*[,1] David R. Beier,† Reinhard Fässler,‡ Dick Heinegård,* and Yngve Sommarin*[,2]

*Department of Cell and Molecular Biology, Section for Connective Tissue Biology, Lund University, S-221 00 Lund, Sweden; †Genetics Division, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; and ‡Max Planck Institute for Biochemistry, Martinsried, Germany

Received June 20, 1997; accepted October 21, 1997

The mouse chondroadherin gene was isolated from a cosmid genomic library by the use of a rat chondroadherin cDNA probe. Southern blot analysis of mouse genomic DNA revealed a simple pattern of hybridization indicating a single copy gene for chondroadherin. The mouse chondroadherin gene encompasses 4.1 kb and consists of four exons separated by one large intron of 1929 bp followed by two smaller introns of 247 and 225 bp, respectively. Most of the translated region, including the start codon and the main part of a leucine-rich region, is contained within the first exon. Two small exons of 164 and 146 bp encode the rest of the protein. Interestingly, 4 bases from the stop codon, in the 3′-UTR, a third intron is located. A putative promoter region of 669 bp was sequenced and shown to contain a potential TATAA-box signal 29 bp upstream of the transcription start site and several recognition sites for transcription factors. The exon/intron organization of the chondroadherin gene differs from those of the other known genes of the leucine-rich repeat (LRR) family in the extracellular matrix. Taken together with comparison of protein sequences of other members of the LRR family in the extracellular matrix, the data suggest that chondroadherin has evolved along a different pathway. The chondroadherin gene was mapped to mouse chromosome 11, near D11Mit14, by single-strand conformation polymorphism linkage analysis.

INTRODUCTION

Cartilage extracellular matrix is dominated by the presence of collagen fibers and the large aggregating proteoglycan, aggrecan. Collagen and aggrecan together are the main molecules responsible for the mechanical properties of cartilage. However, cartilage extracellular matrix also contains a variety of other extracellular matrix proteins. Several of these proteins appear to have roles in the regulation of the assembly of the matrix, e.g. of collagens to fibers, such as decorin (Vogel *et al.*, 1984) or fibromodulin (Hedbom and Heinegård, 1989) and of proteoglycans to aggregates, such as the link protein (Heinegård and Oldberg,1989). Another function of extracellular matrix proteins is to mediate binding of cells. Chondroadherin is a cell-binding protein that has been shown to bind chondrocytes and fibroblasts with similar affinity as fibronectin or collagen (Sommarin *et al.*, 1989). This binding is mediated by an integrin, $\alpha_2\beta_1$ (Camper *et al.*, 1997). Chondroadherin could therefore have important roles in signaling information on matrix properties and function to the cell.

The complete cDNA sequence for bovine chondroadherin has been deduced (Neame *et al.*, 1994), showing that chondroadherin belongs to the family of relatively small, leucine-rich proteins (Patthy, 1987) that are present in the extracellular matrix of cartilage. This family includes biglycan (Fisher *et al.*, 1989), decorin (Day *et al.*, 1986), fibromodulin (Oldberg *et al.*, 1989), lumican (Blochberger *et al.*, 1992), keratocan (Corpuz *et al.*, 1996), PRELP (Bengtsson *et al.*, 1995), proteoglycan-Lb (Shinomura and Kimata, 1992), and osteoinductive factor (Madisen *et al.*, 1990). These proteins are similar in their overall primary structure and have a central region with some 5–11 leucine-rich repeats (LRR) flanked by two disulfide loop regions.

Most of the LRR proteins in the extracellular matrix characterized to date are proteoglycans with one or a few glycosaminoglycan (GAG) chains. Exceptions are PRELP and chondroadherin, which are not substituted with GAGs (Bengtsson *et al.*, 1995; Neame *et al.*, 1994). Both decorin and biglycan can, however, occur without GAG substitution (Johnstone *et al.*, 1993; Roughley *et al.*, 1993). All the LRR proteins, with the exception of chondroadherin, have two or more sites for N-linked

Sequence data from this article have been deposited with the GenBank/EMBL Data Libraries under Accession No. U96626 and with the Mouse Genome Database under Accession No. MGD-JNUM-42830.

[1] To whom correspondence should be addressed at the Department of Cell and Molecular Biology, Section for Connective Tissue Biology, P.O. Box 94, S-221 00 Lund, Sweden. Telephone: +46-46 222 3310. Fax: +46-46 211 34 17. E-mail: Christina.Landgren@medkem.lu.se.

[2] Present address: Wieslab AB, Lund, Sweden.

0888-7543/98 $25.00

oligosaccharide substitution. In some cases these substituents actually may become extended to keratan sulfate chains. Chondroadherin appears to be devoid of carbohydrate substitution except for a small as yet uncharacterized oligosaccharide positioned on Ser^{123} (Neame *et al.*, 1994).

We have isolated the mouse chondroadherin gene to elucidate its genomic organization and to determine similarity with the other known genes of the LRR family. Furthermore, we have determined the chromosomal localization of the gene. The results show that the genomic organization and chromosomal localization differ from the other members of the family. This further strengthens the view that chondroadherin is of a different developmental origin.

MATERIALS AND METHODS

Southern blot analysis. Mouse genomic DNA was isolated from liver tissue as described (Ausubel *et al.*, 1994) and initially characterized by Southern blot hybridization. Ten micrograms of DNA was digested with 2 × 10 units of restriction enzymes *Bgl*II, *Hin*dIII, and *Kpn*I (Life Technology) for 2 × 30 min at 37°C. Fragments were separated on 0.8% agarose gel, transferred onto positively charged nylon membrane (Hybond N^+; Amersham), and hybridized according to Church and Gilbert (1984). The hybridization was performed using an 887-bp *Nco*I–*Sca*I fragment from a full-length rat cDNA clone (Shen *et al.*, 1998) (GenBank Accession No. AF004953) as the probe (nucleotide +62 to +949 in the complete cDNA sequence of rat chondroadherin). Radiolabel was detected by exposure of X-ray film at −70°C in cassettes with intensifying screens.

Isolation of a genomic clone. A mouse genomic cosmid library was screened using the 887-bp *Nco*I–*Sca*I rat cDNA fragment as the probe. The probe included the ATG start codon. After two rounds of screening one positive clone was selected. The DNA corresponding to the chondroadherin gene was held within a cosmid referred to as sCos-I (Evans *et al.*, 1989).

Characterization of the cosmid clone. The cosmid clone was analyzed by digestion of 2 μg of cosmid DNA with the restriction enzymes *Bam*HI, *Sac*I, *Hin*dIII, and *Eco*RI (Life Technology). The fragments were separated and hybridized as above using rat cDNA probes for the 5′ end (*Eco*RI–*Stu*I, 338 bp), the 3′ end (*Nco*I–*Eco*RI, 285 bp), and an *Eco*RI-digested full-length rat cDNA clone (1673 bp).

DNA sequencing. Genomic fragments from the cosmid clone were subcloned into pBluescript KS II. The fragments of 2500 and 4000 bp, respectively, were sequenced by the ABI Prism Dye Terminator cycle sequencing kit (Applied Biosystems, Perkin–Elmer), first using T3 and T7 primers and internal oligonucleotide primers (18- to 21-mers) synthesized from the rat cDNA sequence and later from the new mouse DNA sequence.

The sequencing reaction products were analyzed on an automated sequencer apparatus Model 373 A (Applied Biosystems, Perkin–Elmer). Exons were sequenced in both directions. Introns were sequenced in one direction only, except for regions with poor reliability, of which both strands were sequenced. Analyses of sequences were performed using the PC Gene (Intelligenetics) program package.

Single-strand conformation PCR analysis. Primers were designed to amplify a region corresponding to intronic sequences of chondroadherin to test for single-strand conformation polymorphisms (SSCPs) between mouse strains. These were analyzed as previously described (Beier, 1993). Briefly, oligonucleotides were radiolabeled with [^{32}P]ATP using polynucleotide kinase, and genomic DNAs from a series of mouse strains were amplified using standard protocols (anneal 55°C for 1 min, extend at 72°C for 2 min, and denature at 94°C for 1 min for 40 cycles, with a final extension at 72°C). Two microliters of the amplified reaction mixture was then added to 8.5 ml USB (United States Biochemical Corp.) stop solution, denatured at 94°C for 5 min, and immediately placed on ice. Two microliters of each reaction mixture was then loaded onto a 6% nondenaturing acrylamide sequencing gel and electrophoresed in 0.5× TBE buffer for 2–3 h at 40 W in a 4°C cold room. A primer pair with the sequence ACGAAGGCTGATTTAGAATGAGG (forward) and GTATTGGTGCCCTCCTCTGAG (reverse) identified a polymorphism between C57BL/6J and *Mus spretus* and was used to analyze DNA prepared from the BSS backcross (Rowe *et al.*, 1994). The selected primer pair amplified nucleotides 2606 to 2850 in intron 1. The identity of the amplified fragment was confirmed by cleavage with *Acc*I restriction enzyme. The allele distribution pattern was analyzed using the Map Manager Program (Manly, 1993).

Restriction mapping. Restriction mapping of the cosmid clone was done essentially as described by Evans *et al.* (1989).

Primer extension. A 21-base oligonucleotide, 5′-AGACCAGACTGAATAAGAGCG, corresponding to the reverse complement from position +60 to +80 in the first exon of the mouse sequence was end labeled with [γ-^{32}P]dATP. Forty micrograms of total RNA isolated from mouse trachea was mixed with 10^6 cpm of labeled oligonucleotide in 0.15 M KCl, 0.01 M Tris–HCl, pH 8.3, 1 mM EDTA, 65°C for 90 min. Reverse transcription was performed using Superscript II (Gibco BRL) according to *Current Protocols in Molecular Biology*. After RNase digestion the product was phenol extracted, ethanol precipitated, and dissolved in loading buffer. The sample was separated on a 6% polyacrylamide gel. A dideoxy sequencing reaction of genomic cosmid mouse DNA primed with the same oligonucleotide was used as the standard.

Reverse transcriptase PCR. Three micrograms of total RNA isolated from mouse trachea was used to synthesize cDNA using 2 pmol of specific primer, 5′-CAGCGCTGTGCATCCGCA, corresponding to the reverse complement from position 3621 to 3638 in the fourth exon; 200 ng oligo(dT); and Superscript II (Gibco BRL) under conditions recommended by the manufacturer. A small aliquot, 2 μl, of this reaction was used for PCR amplification using *Taq* DNA polymerase (Gibco BRL) and the following primers flanking the introns: intron 1, 5′-ACCTGCGCTGGCTCTACCTGT and 5′-GGTCTCCAGGTTGTCAAA; intron 2, 5′-CTCAGATGCTGCCTTCTC and 5′-CATCTGTGTCACGAATCC; and intron 3, 5′-CTCGCCAGCCAAGTTCAA and 5′-GAGGCTGTAGGAGAAGGTGTG.

To obtain information on intron size, PCR amplification of genomic DNA isolated from mouse liver was done using standard methods as described (*Current Protocols*) and the same primers as above. The resulting PCR products were analyzed on a 0.8% agarose gel.

RESULTS

Analysis of mouse genomic DNA. Southern blot analysis was performed on mouse genomic DNA. The blot was hybridized with a rat chondroadherin cDNA fragment. A simple pattern of hybridization consistent with the presence of a single-copy gene for chondroadherin was seen (Fig. 1).

Analysis of the genomic clone. A mouse genomic library was screened using the same rat probe as above. The screening of the library yielded one clone containing the whole chondroadherin gene. The DNA of the genomic clone was digested to completion with the restriction enzymes *Bam*HI, *Sac*I, *Eco*RI, and *Hin*dIII. The size of the entire insert in the cosmid clone was estimated at 31 kb by sizing the fragments on an agarose gel. A crude map of the coding regions was constructed using rat cDNA probes in Southern hybridization analyses. This experiment revealed two *Sac*I fragments of about 2.5 and 4 kb encoding the 5′ and 3′ region of the cDNA. Southern hybridization using a

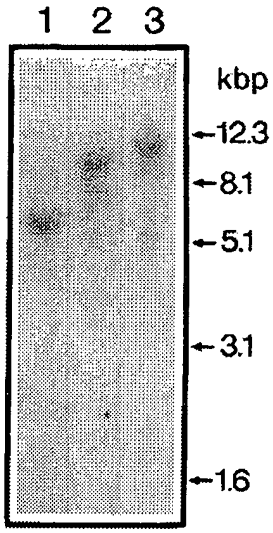

FIG. 1. Detection of the chondroadherin gene in mouse genomic DNA. Southern blot analysis was performed with 10 μg each of mouse genomic DNA digested with restriction enzymes as indicated, probing with an 887-bp rat chondroadherin cDNA fragment. The DNA was digested with the following enzymes: Lane **1**, *Bgl*II; lane **2**, *Hin*dIII; and lane **3**, *Kpn*I. A simple pattern of hybridization is seen, suggesting a single-copy gene for chondroadherin.

full-length rat cDNA clone gave the same hybridization pattern (not shown), indicating that the mouse chondroadherin gene was completely included in this 6.5 kb of genomic DNA.

Structure of the gene. The two *Sac*I fragments of the genomic clone were subcloned into pBluescript and sequenced in total, shown in Fig. 2. The results show that the mouse chondroadherin gene consists of four exons spanning 4.1 kb of genomic DNA, not including the putative promoter region, schematically shown in Fig. 3. The intron/exon organization divides the coding sequence of chondroadherin into three exons. The first exon encodes most of the coding region including the 5'-UTR, the ATG translational start, and 9 of the 11 leucine-rich repeats. The second and third exons encode the remainder of the coding region, and these exons divide the protein sequence such that the 4 carboxy-terminal cysteines are located in two exons. The fourth exon is entirely held within the 3'-UTR and contains the putative polyadenylation signal. The first intron is comparatively large, 1929 bp long, and splits the gene in a location corresponding to the ninth leucine-rich repeat. This first intron also contains a twice-repeated, 48-nucleotide element at position 2316 to 2363 and at position 2370 to 2417. Only 1 nucleotide differs between the two sequences. The ninth leucine-rich repeat is spliced between residues 14 and 15, indicating a phase 0 intron. The second intron, 247 bp long, splices at the 23rd amino acid residue in the last repeat after the second nucleotide, phase II. Sequence analysis indicated the presence of a third intron located only 4 bases from the translation termination codon in the 3'-UTR. Alignment of rat or bovine cDNA with the mouse gene sequence showed high similarity in the 3'-UTR except for a potential 225-bp large insertion that has no similarity to the rat or bovine cDNA sequence (not shown). Furthermore, the ends of this potential intron have the GT/AG sequence typical for introns. To confirm that all potential introns were spliced out in mature mouse mRNA, RT-PCR was performed using mouse tracheal RNA. The cDNA was used to amplify fragments using primers flanking the introns. The sizes of the products were compared to the sizes of products from PCR amplification using the same primers but with genomic DNA as a template. This analysis confirmed that all the putative introns were spliced out in the fragments amplified from the cDNA (Fig. 4). Additionally, all introns show the classical GT/AG sequence flanking the intron splice junctions in the 5' and 3' ends of the introns (Breathnach and Chambon, 1981) (Fig. 2).

The mouse nucleotide sequence of the translated regions shows a homology of 94%, compared with the cDNA sequence of rat chondroadherin (Shen *et al.*, 1998), and of 86% with bovine chondroadherin (Neame *et al.*, 1994).

Polyadenylation signal. Sequence determination of 2665 bp of genomic sequence downstream of the TAA stop codon to the end of the second *Sac*II fragment did not reveal a classical AATAAA polyadenylation signal. The best candidate for a polyadenylation signal is the repeat TATAAACATAAA found at position 4047, underlined in Fig. 2. This sequence is also found at the 3' end of the 1644-bases-long, full-length rat cDNA clone just upstream of a poly(A) stretch (Shen *et al.*, 1998). This putative polyadenylation signal gives a total length of the mature mRNA of 1.6 kb. This agrees well with the size of mRNA found on Northern blots of bovine (Neame *et al.*, 1994), rat, and mouse chondroadherin mRNA (data not shown).

Transcription start site. A primer extension analysis was done with a primer complementary to nucleotides +60 to +80 just downstream of the ATG translational start codon. The analysis places the transcrip-

FIG. 2. DNA sequence and deduced amino acid sequence of the mouse chondroadherin gene. The sequence shown begins with the putative promoter region and ends with the probable 3' end of the mRNA. A total of 669 bp of genomic DNA upstream of the ATG was sequenced and analyzed with the PC Gene (Intelligenetics) program package and the GCG program package. A TATAA-box signal (double underline) was found at −29 bp upstream of the putative transcription start site, identified by primer extension analysis (numbered +1). Several recognition sites for transcription factors were found as indicated (bold, italics) (GATA, globin-activating site). The sequences of exons 1–4 are written in uppercase letters. Intron 1 is truncated as shown with a double-slash. Intron sequences including the gt/ag sequences at the boundaries are written in lowercase letters. Signal peptide sequence is in italic, cysteine residues are double underlined, and the position of Ser^{123}, the only site at which a posttranslational modification has been found, is indicated by a filled triangle. The positions of the LRRs are indicated. The atypical putative polyadenylation signal is single underlined.

```
-669 bp
GAGCTCTTACGGCCTGGTGCCACTGGGCTCCGAGAAGGGGCAGAGCCAAACGCACGGCTGTCAGCTAGCCT   -598
CTGCAACCAGCTCCCCCACCTCCTTGGGATAAACTGAGGAACCCAGAAGCGGGAGCCCAACCCACAGCAGCT  -526
CTCACGCTCCGCCTGCGCCGCACAACAGTCCCATTAAAGCGCCGCCGGCTGGCCGACCGCGGTGAGACGCAT  -454
CCGGCTGTCGGGCCCCACTTCCTCCCTCCCGGAGTCCAGGGTGACCTGTCTGCCAAGGGTGTATGGGGGAAG  -382
                                                                c-Ets
GAGACGTAGAGAACTCAAACTTGAGCAAATAAATAAGTTCTGGGAACACTTCCCTCTGCCCAGTGGAAATTC  -310
            GATA-1
AGAAGCCCCTCGACACACCTATCACCGTCCACCCCACCTCGGGGTGTTGGTCCAGATAGAGGAGGGTAGGGG  -238
                    c-Ets
AAGGTGCAGCATAATGTTTGCAAACAGGAACCAAGGGGTTGGGGTTCAGGGGAAGGGCCCTCAGCCCTACAC  -166
                                       GATA-1              Sp1
ACGGTCTCCTGCTGTGAAAAGAGGCCCCCAGCCATCGAGGATGGGGAAGCATCTCTGGGCGGGAAGGGTTAA   -94
                                                                -29 bp
ATCAGTGGCTTCGGTGCTCCACGTAGTAGCTGGCTCCGCTGCCAACTGCGGTCAAGGCTGCCCTATAAATGG   -22
GCCGGGAGACCCGAGAGTCGA                                                       -1
+1 bp
GGACTTGTCGCTGCCTTAGCCCCCAGCCCAGGCTCAAGGCGTTCTAACCATGGCCCGCGCGCTCTTATTCAGT   73
                                                  M  A  R  A  L  L  F  S   -13

CTGGTCTTTCTTGCCATCCTCCTGCCTGCGCTAGCCGCCTGCCCCCAAAACTGCCACTGCCATGGAGATCTG  145
 L  V  F  L  A  I  L  L  P  A  L  A  A  C  P  Q  N  C  H  C  H  G  D  L   12

CAGCATGTCATCTGCGACAAGGTGGGGCTGCAGAAGATCCCCAAGGTATCAGAGACAACCAAACTGCTCAAT  217
 Q  H  V  I  C  D  K  V  G  L  Q  K  I  P  K  V  S  E |T  T  K  L  L  N   36
                                                       LRR 1
CTCCAGCGCAACAACTTCCCGGTGCTGGCTGCCAACTCGTTTCGGACCATGCCGAACCTGGTCTCCCTGCAC  289
 L  Q  R  N  N  F  P  V  L  A  A  N  S  F  R  T  M  P |N  L  V  S  L  H   60
                                                       LRR 2

CTGCAACACTGCAACATCCGCGAGGTGGCGGCTGGTGCCTTCCGAGGCCTGAAGCAGCTTATCTACCTGTAC  361
 L  Q  H  C  N  I  R  E  V  A  A  G  A  F  R  G  L  K |Q  L  I  Y  L  Y   84
                                                       LRR 3

CTGTCCCACAACGACATCCGGGTATTGCGAGCTGGAGCCTTCGACGACCTGACTGAACTCACTTACCTCTAT  433
 L  S  H  N  D  I  R  V  L  R  A  G  A  F  D  D  L  T |E  L  T  Y  L  Y  108
                                                       LRR 4

CTAGACCACAACAAAGTGTCGGAACTGCCCCGGGGGTTGCTCTCTCCTCTGGTCAACCTCTTCATCTTGCAA  505
 L  D  H  N  K  V  S  E  L  P  R  G  L  L  S  P  L  V |N  L  F  I  L  Q  132
                                              ▲        LRR 5

CTCAACAACAACAAAATCCGAGAGCTGCGTGCTGGAGCTTTCCAGGGGGCCAAGGACCTGCGCTGGCTCTAC  577
 L  N  N  N  K  I  R  E  L  R  A  G  A  F  Q  G  A  K |D  L  R  W  L  Y  156
                                                       LRR 6

CTGTCAGAAAATGCCCTCAGTTCCCTGCAGCCTGGTTCCCTGGATGATGTGGAGAACCTAGCCAAGTTCCAC  649
 L  S  E  N  A  L  S  S  L  Q  P  G  S  L  D  D  V  E |N  L  A  K  F  H  180
                                                       LRR 7

CTGGACAAGAACCAGCTGTCTAGCTACCCCTCAGCCGCCCTGAGCAAACTTCGGGTGGTGGAGGAGCTGAAG  721
 L  D  K  N  Q  L  S  S  Y  P  S  A  A  L  S  K  L  R |V  V  E  E  L  K  204
                                                       LRR 8

CTGTCTCACAACCCTCTGAAGAGCATCCCAGACAATGCCTTCCAGTCCTTCGGTAGATATCTGGAGACCCTC  793
 L  S  H  N  P  L  K  S  I  P  D  N  A  F  Q  S  F  G  R |Y  L  E  T  L  228
                                                          LRR 9

TGGCTGGATAACACCAACCTGGAGAAGgtaagtgccccagctgcagtt/  /ctgcctccctcacctcacag 2749
 W  L  D  N  T  N  L  E  K                                                237

TTCTCAGATGCTGCCTTCTCGGGTGTGACCACACTGAAACACGTCCATCTGGACAACAACCGCCTGAACCAA 2821
 F  S  D  A  A  F  S  G  V  T |T  L  K  H  V  H  L  D  N  N  R  L  N  Q  261
                               LRR 10

CTGCCTTCCTCCTTCCCCTTTGACAACCTGGAGACCCTCACTCTCACCAACAACCCATGGAAATGCACCTGC 2893
 L  P  S  S  F  P  P  F  D |N  L  E  T  L  T  L  T  N  N  P  W  K  C  T  C  285
                            LRR 11

CAGCTCCGTGGCCTTCGGCGgtgagaatattcctccatataacccccagactgccgtccacatgacagacgg 2965
 Q  L  R  G  L  R  R                                                      292

tcctagagtaggacagcctggacatcctagtcagctacctagcatgtcgggtactgagtggttcccttctct 3037
catttgtcaaatgaagatgacaactccagatatttctatggccatagtccatcccggtcactgtccctttcc 3109

caagccttcccacccagcttttccaagcccagcaactctttgtctctgtagGTGGTTGGAAGCCAAGGCTT   3180
                                                     W  L  E  A  K  A     298

CTCGACCGGATGCTACCTGCTCCTCGCCAGCCAAGTTCAAGGGTCAGCGGATTCGTGACACAGATGCCCTTC 3252
 S  R  P  D  A  T  C  S  S  P  A  K  F  K  G  Q  R  I  R  D  T  D  A  L  322

GCAGCTGCAAATCCCCGACCAAGAGGTCCAAGAAAGCTGGCCGCCATTAAACAGgtgggggctgggtaggga 3324
 R  S  C  K  S  P  T  K  R  S  K  K  A  G  R  H  -                        338

ggccaccacggtctacctttggaaattccagatggggtgctgctatatcccatgacaccacttccggaggag 3396
caatcagttccctgtcttacaagaaaaggagggaggacaggataacctctcccatggcttggcctaggacgt 3468
ccatgggtccctttaatgactctgggtgactggaatcctaatacccatcttctctcactatagGTCCTGATC 3540
CAGCCAGTCCTGGCGACTGCCTTCCGCTGGAGAGACTACTGACGTTCCCTCCCATCATCCACACCTTCTCCT 3612
ACAGCCTCTGCGGATGCACAGCGCTGCCCCGCCCCCGCCCCCACCTAGGTACATCCTGGCAGGGGCACTGGG 3684
CTCTCTATCACCATCCCAGCTCCACCCAGTGGGGTCCTAGGAAAGACACAGAATCCCTCCCCAGCCACTGTG 3756
TCTGGGCTCTGCCATGGCTCCTTTGAGAGAAGCTATTGTAGAACCTCCTACCCTCTGTCCATCGGAGCTAAA 3828
GCGCAGTGGTCATTGGGATGACCACGTTATTACCACCTTCCTCGGTTCCCTCTGTCCCTGCCATTTGGAAAC 3900
AAACATCAGGCCCCTGACCCACCCTGATTGCCAGAAAGAATTTCAGGCCCATGCCCCAACTCTGCCAGTTCC 3972
TGCCTGCCAGGACATGCTACCAGGATACCAGTAGCGCTTGGCTGCATATCCTTCCTGTTTGCGCTCCAGATT 4044
TCTATAAACATAAATGTATGTGTGTTCA                                              4072
```

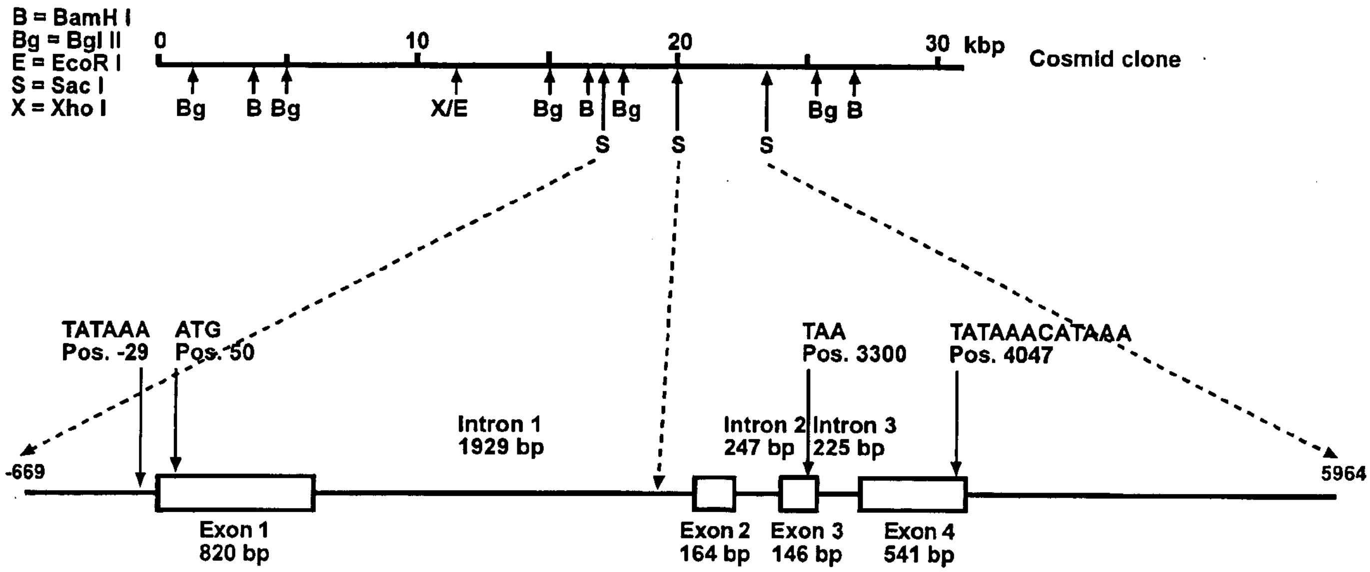

FIG. 3. Organization of the mouse chondroadherin gene. A restriction map of the cosmid clone is shown at the top. The structures of the two *Sac*I fragments that were sequenced in full are shown schematically. The gene spans about 4.1 kb of genomic DNA. The boxes show the locations of the exons. Approximate locations of the putative TATAA-box signal and the ATG translational start site are indicated. Exon 1 encompasses the main part of the translated region. It is separated from the next exon by a 1929-bp intron. This is followed by two smaller exons interrupted by a 247-bp intron. The stop signal is found in the third exon as indicated. The third intron is located in the 3′-UTR. A putative polyadenylation signal is found in the fourth exon.

tional start site, beginning with GGAC, 50 bp upstream of the ATG (Fig. 5). This site coordinates with the start site of the cDNA for rat and bovine chondroadherin, which are both very similar to the mouse sequence. The rat and bovine full-length cDNAs both start with the sequence GGAC as in the mouse genomic DNA sequence.

Protein sequence. The protein-coding sequence of mouse chondroadherin corresponds to 358 amino acids, including the signal peptide of 20 amino acids. Alignment of the amino acid sequence of mouse chondroadherin with rat and bovine chondroadherin shows similarities of 97 and 92%, respectively. The leucines, cysteines, and other hydrophobic residues in the LRR core region are all conserved in the three species. The only position with a posttranslational substitution, i.e., Ser^{123}, is conserved among all species.

Putative promoter region. A total of 669 bp of genomic DNA upstream of the transcriptional start site was sequenced and analyzed for the presence of sequences typical of mammalian promoters. A TATAA-box sequence was found 29 bp upstream of the transcriptional start site (Fig. 2). This distance from the transcriptional start site agrees well with that of other known TATAA boxes. The analysis did not reveal any consensus sequence for a CCAAT box in this region. Other possible regions upstream of this region all showed lower scores in the searches. Upstream of the TATAA-box several potential recognition sites for transcription factors were found as indicated in Fig. 2.

Chromosome localization. SSCP was used to map chondroadherin (Beier, 1993; Beier *et al.*, 1992). Primers corresponding to intronic sequence of chondroadherin were analyzed and identified an SSCP between inbred mouse strains (see Material and Methods). The BSS interspecific backcross was genotyped and the allele distribution pattern analyzed using the Map Manager program. Chondroadherin mapped to chromosome 11 with a lod likelihood score of 24.7. Three genotypes generated double-crossovers with adjacent loci and were not included in the analysis since they are likely to be incorrectly typed. No recombinants were found between chondroadherin and a cluster of genes that includes Hoxb7, Hoxb13, Coxa5, Nrdf, and Csfg, as well as the microsatellite marker D11Mit14. The position of chondroadherin with respect to flanking microsatellite markers is D11Mit36–4.4 ± 2.2 cM–chondroadherin–3.3 ± 1.9 cM–D11Mit10.

DISCUSSION

We have isolated and sequenced the gene for mouse chondroadherin, characterized the intron/exon organization, and determined the transcriptional start site and the chromosomal localization of the gene. Mouse chondroadherin was shown to represent a single-copy gene, and it is spread over about 4.1 kb of genomic DNA. This makes the chondroadherin gene smaller than the other known genes of the structurally related extracellular matrix LRR proteins, which range from 7.5 kb for the human lumican gene to 38 kb for the human decorin gene. About half the number of genes encoding the different LRR proteins have been characterized (Kobe and Deisenhofer, 1994). A comparison of the primary structure of LRR proteins, for which the sequences are known, shows that decorin and biglycan form one class, whereas lumican, fibromodulin, and

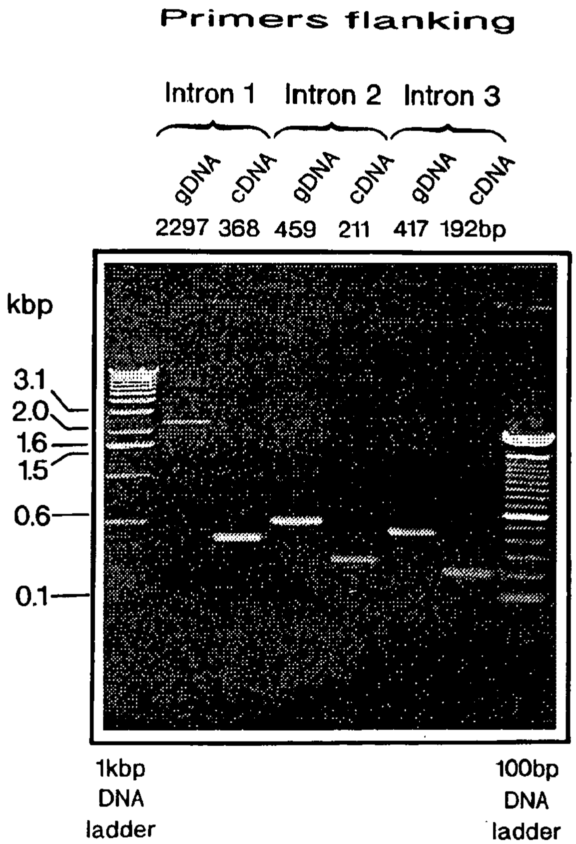

FIG. 4. Verification of intron localization by PCR. The locations of introns were verified by comparison of PCR on genomic DNA and tracheal cartilage cDNA using primers flanking sides of possible introns. The sizes of the predicted fragments are shown at the top.

PRELP form a second class and chondroadherin a third class (Bengtsson *et al.*, 1995). This relationship also holds at the gene structure level. The decorin (Scholzen *et al.*, 1994) and biglycan (Wegrowski *et al.*, 1995) genes are encoded by eight exons with the LRRs in six exons. The fibromodulin (Antonsson *et al.*, 1993), lumican (Grover *et al.*, 1995), and PRELP (Grover *et al.*, 1996) genes are composed of three exons with the LRRs encoded by a single exon. The chondroadherin gene differs by having a total of four exons with the LRRs encoded by two exons. The first intron splits the ninth LRR between residues 14 and 15 in the putative α-helix in a phase 0 manner. The second intron splices the 11th repeat between nucleotides 2 and 3 in the 23rd residue, indicating a phase 2 intron, so also in the putative α-helix. The third intron is located only 4 bases from the termination codon. In the decorin and biglycan genes the introns often localize at similar positions, most frequently between residues 1 and 6 in the β-sheet region of the LRR. This position is the most frequent intron localization among LRR proteins (Kobe and Deisenhofer, 1994), with chondroadherin being an exception to this general pattern.

The consensus sequence of the LRR is xLxxLxLxx-[N,C,T]x[L,I]xxaP followed by 5 to 10 residues. Two types of repeats are recognized that differ at position 10 in the consensus sequence. One is the A-type repeat with a cysteine/threonine and the other the B-type with an aspargine at this position (Kobe and Deisenhofer, 1994). The length of the LRRs varies within the family. The LRRs tend to appear in triplets with two longer repeats followed by a short repeat (Bengtsson *et al.*, 1995). Biglycan, decorin, fibromodulin, lumican, and PRELP all have exclusively B-type repeats. These vary in length from 20 to 26 amino acids. Again chondroadherin is different with the 2nd and 9th LRR being A-type repeats with a cysteine and a threonine, respectively. In addition, chondroadherin shows a distinct pattern with LRRs of equal length, being composed of 24 residues. Exceptions are the 8th repeat, which contains 25 residues, and the 10th repeat, which appears to contain only 22 residues.

The cysteine pattern in the C-terminal region of chondroadherin is different from those of the other extracellular matrix LRR proteins, with four cysteine residues being divided into two exons in a way that the disulfide bonds are made between cysteine residues in different exons. The other genes have only two cysteine residues in this region. These cysteines are encoded by a single exon separate from the LRRs. The C-terminal cysteine pattern found in chondroadherin is also found in the platelet glycoproteins V (Lanza *et al.*, 1993), IX (Hickey *et al.*, 1990), 1b-α (Lopez *et al.*, 1987), and 1b-β (Wicki *et al.*, 1989). The genomic structures of some of these genes are known (Yagi *et al.*, 1995). The exon/intron organization of these genes differs from the chondroadherin gene in that the entire protein coding sequence is within one exon. Thus, the chondroadherin C-terminal disulfide bonded region is likely to have been created by recombination events different from those of the other known genes.

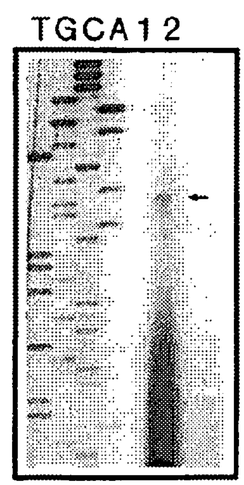

FIG. 5. Primer extension analysis of the mouse chondroadherin gene places the transcriptional start 50 bp upstream of the ATG. An end-labeled oligonucleotide corresponding to the inversed complement from position +60 to +80, in the first exon of the mouse sequence, was annealed to total RNA isolated from mouse trachea. Primer extension was performed as described under Materials and Methods. Lanes T, G, C, and A correspond to a dideoxy sequencing reaction of genomic DNA primed with the same oligonucleotide. Lane 1, blank; lane 2, a major band corresponding to the transcriptional start site is indicated by an arrow.

```
MURINE   CTTCCTGT-TTGC-GCTCCAGATTTCTATAAACATAAATGTATGTGTGTTCA         50
RAT      CATGCTGTATTTCTGCCCCGGATTTCTATAAACATAAATGTCTGTGTGTAAAAAAAAAA  59
BOVINE   CACGCTGCATTTCTTCCCCAGATTTCTATAAATATAAATTTATGTATGTATAATAA     56
         *   ***  ** *  * ** ************ ****** * *** ***  *
```

FIG. 6. Alignment of the 3′ end nucleotide sequence of chondroadherin cDNA and genomic DNA from different species. Comparison of the 3′ end nucleotide sequence of murine chondroadherin with those of rat and bovine chondroadherin is shown. All three species have similar sequences, and all contain the atypical polyadenylation signal (double underlined). Identical nucleotides are indicated by stars.

The presence of the third intron in the 3′-UTR also sets chondroadherin apart compared to the others. Additionally, all the other LRR protein genes mentioned above, including the platelet glycoproteins, have an intron close to the ATG translational start codon in the 5′ end, which is not present in the chondroadherin gene.

About 669 bp of a putative promoter region was sequenced and examined to determine if any specific structural features of mammalian promoters were present. By sequence analysis we found a potential TATAA-box sequence situated 29 bp upstream of the putative transcription initiation site identified by primer extension analysis. Such a TATAA box is often found in proteins that have a relatively restricted expression, which agrees well with the rather tissue-specific expression pattern of chondroadherin (Larsson *et al.*, 1991). Decorin has been shown to possess a TATAA-box sequence, while this is absent in biglycan. It has been suggested that TATAA-less promoters are common for proteins with a more widespread expression (Ishii *et al.*, 1985).

It is not unambiguous to pinpoint the end of the fourth putative exon since no classical polyadenylation signal was found from the stop codon to the end of the second *Sac*I fragment. The best candidate for a polyadenylation signal is the repeat TATAAACATAAA found at position 4047, underlined in Fig. 2. This sequence is also found at the 3′ end of the 1644-bases-long, full-length rat cDNA clone just upstream of a poly(A) stretch (Shen *et al.*, 1998) and in the 3′ end of the bovine cDNA clone (Fig. 6). It appears, then, that the mouse, bovine, and rat genes all have atypical polyadenylation signals. This putative polyadenylation signal differs from the classical consensus signal by 1 nucleotide. Such a naturally mutated polyadenylation signal may be weaker than normal. There is some evidence that this type of site has stronger upstream and downstream elements that are involved in defining the cleavage site to compensate for the weaker signals (for reference see Wahle, 1995). Such elements are not fully characterized, but the downstream elements seem to occur within 50 bases 3′ of the cleavage site. One form of such an element may have the consensus sequence YGUGUUYY (Y = C or T) (Wahle, 1995). The sequence TGTGTTCA is found 17 bp downstream of the putative polyadenylation site in the mouse genomic sequence. A similar sequence is present also in the rat cDNA sequence.

The primary structure of mouse chondroadherin shows a high degree of conservation when compared to the same protein from other species. The identity of the mouse chondroadherin amino acid sequence was more than 90% when compared to rat and bovine chondroadherin amino acid sequences. Furthermore, the cysteine pattern and the leucine residues in the LRRs were conserved in all three species. Also, the other members of the LRR family show a high overall homology between species, indicating important well-conserved biological functions.

The chromosomal localizations of these genes do not follow their similarity in protein structure. Thus lumican and decorin are both found on mouse chromosome 10, while fibromodulin and PRELP are both found on human chromosome 1. Biglycan, which is most similar to decorin, is found on human chromosome X. SSCP analysis was used to localize chondroadherin to mouse chromosome 11, tightly linked to the Hoxb cluster. This locus has been mapped to 17q21–q22 in humans. Since subchromosomal linkage relationships are conserved in many cases between mouse and human, this result suggests that the human homolog of chondroadherin will also be found in this region. In addition, given the well-described paralogy between this region on mouse chromosome 11 and the Hoxc region on mouse chromosome 15 (corresponding to human 12q13), it is quite possible that there is another chondroadherin-like gene in mammals. Several murine mutations that, according to the Mouse Genome Database (www.informatics.jax.org), map close to the position of chondroadherin on chromosome 11 and whose pathology could be due to abnormalities of extracellular matrix development include Bald arthritic (Bda), Bareskin (Bsk), Rex (Re), and Cleft lip1 (clfl).

ACKNOWLEDGMENTS

This work was supported by the Medical Faculty, Lund University, the Swedish Medical Research Council, Margaret and Axel Ax:son Johnson's Stiftelse, Greta and Johan Kock's Stiftelser, and Konung Gustaf V:s 80-årsfond. C.L. was on a graduate training grant from the Medical Faculty, Lund University.

REFERENCES

Antonsson, P., Heinegård, D., and Oldberg, Å. (1993). Structure and deduced amino acid sequence of the human fibromodulin gene. *Biochim. Biophys. Acta* **1174:** 204–206.

Ausubel, B., *et al.*, (1994). "Current Protocols in Molecular Biology", Wiley, New York.

Beier, D.R. (1993). Single-strand conformation polymorphism (SSCP) analysis as a tool for genetic mapping. *Mamm. Genome* **4:** 627–631.

Beier, D. R., Dushkin, H., and Sussman, D. J. (1992). Mapping genes in the mouse using single-strand conformation polymorphism anal-

ysis of recombinant inbred strains and interspecific crosses. *Proc. Natl. Acad. Sci. USA* **89:** 9102–9106.

Bengtsson, E., Neame, P. J., Heinegård, D., and Sommarin, Y. (1995). The primary structure of a basic leucine-rich repeat protein, PRELP, found in connective tissues. *J. Biol. Chem.* **270:** 25639–25644.

Blochberger, T. C., Vergnes, J-P., Hempel, J., and Hassell, J. R. (1992). cDNA to chick lumican (corneal keratan sulfate proteoglycan) reveals homology to the small interstitial proteoglycan gene family and expression in muscle and intestine. *J. Biol. Chem.* **267:** 347–352.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50:** 349–383.

Camper, L., Heinegård, D., and Lundgren-Åkerlund, Evy. (1997). Integrin $\alpha_2\beta_1$ is a receptor for the cartilage matrix protein chondroadherin. *J. Cell Biol.* **138:** 1159–1167.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81:** 1991–1995.

Corpuz, L. M., Funderburgh, J. L., Funderburgh, M. L., Bottomley, G. S., Prakash, S., and Conrad, G. W. (1996). Molecular cloning and tissue distribution of keratocan. *J. Biol. Chem.* **271:** 9759–9763.

Day, A. A., Ramis, C. I., Fisher, L. W., Gehron-Robey, P., Termine, J. D., and Young, M. F. (1986). Characterization of bone PG II cDNA and its relationship to PG II mRNA from other connective tissues. *Nucleic Acids Res.* **14:** 9861–9876.

Evans, G. A., Lewis, K., and Rothenberg, B. E. (1989). High efficiency vectors for cosmid microcloning and genomic analysis. *Gene* **79:** 9–20.

Fisher, L. W., Termine, J. D., and Young, M. F. (1989). Deduced protein sequence of bone small proteoglycan I (biglycan) shows homology with proteoglycan II (decorin) and several nonconnective tissue proteins in a variety of species. *J. Biol. Chem.* **264:** 4571–4576.

Grover, J., Chen, X-N., Korenberg, J. R., Recklies, A. D., and Roughley, P. J. (1996). The gene organization, chromosome location and expression of a 55-kDa matrix protein (PRELP) of human articular cartilage. *Genomics* **38:** 109–117.

Grover, J., Chen, X-N., Korenberg, J. R., and Roughley, P. J. (1995). The human lumican gene: Organization, chromosomal location and expression in articular cartilage. *J. Biol. Chem.* **270:** 21942–21949.

Hedbom, E., and Heinegård, D. (1989). Interaction of a 59-kDa connective tissue matrix protein with collagen I and collagen II. *J. Biol. Chem.* **264:** 6898–6905.

Heinegård, D., and Oldberg, Å. (1989). Structure and biology of cartilage and bone matrix noncollagenous macromolecules. *FASEB J.* **3:** 2042–2051.

Her, C., and Weinshilboum, R. M. (1995). Rapid restriction mapping by use of long PCR. *BioTechniques* **19:** 530–532.

Hickey, M. J., Deaven, L. L., and Roth, G. J. (1990). Human platelet glycoprotein IX: Characterization of cDNA and localization of the gene to chromosome 3. *FEBS Lett.* **274:** 189–192.

Ishii, S., Merlino, G. T., and Pastan, I. (1985). Promoter region of the human Harvey ras proto-oncogene: Similarity to the EGF receptor proto-oncogene promoter. *Science* **230:** 1378–1381.

Johnstone, B., Markopoulos, M., Neame, P., and Caterson, B. (1993). Identification and characterization of glycanated and non-glycanated forms of biglycan and decorin in the human intervertebral disc. *Biochem. J.* **292:** 661–666.

Kobe, B., and Deisenhofer, J. (1994). The leucine-rich repeat: A versatile binding motif. *Trends Biochem. Sci.* **19:** 415–421.

Lanza, F., Morales, M., de La Salle, C., Cazenave, J. P., Clemetson, K. J., Shinomura, T., and Phillips, D. R. (1993). Cloning and characterization of the gene encoding human platelet glycoprotein V: A member of the leucine-rich glycoprotein family cleaved during thrombin induced platelet activation. *J. Biol. Chem.* **268:** 20801–20807.

Larsson, T., Sommarin, Y., Paulsson, M., Antonsson, P., Hedbom, E., Wendel, M., and Heinegård, D. (1991). Cartilage matrix proteins: A basic 36 kDa protein with a restricted distribution to cartilage and bone. *J. Biol. Chem.* **266:** 20428–20433.

Lopez, J. A., Chung, D. W., Fujikawa, K., Hagen, F. S., Papayannopoulou, T., and Roth, G. J. (1987). Cloning of the alpha chain of human platelet glycoprotein Ib: A transmembrane protein with homology to leucine-rich alpha 2-glycoprotein. *Proc. Natl. Acad. Sci. USA* **84:** 5615–5619.

Madisen, L., Neubauer, M., Plowman, G., Rosen, D., Segarini, P., Dasch, J., Thompson, A., Ziman, J., Bentz, H., and Purchio, A. F. (1990). Molecular cloning of a novel bone-forming compound: Osteoinductive factor. *DNA Cell Biol.* **9:** 303–309.

Manly, K. F. (1993). A Macintosh program for storage and analysis of experimental genetic mapping data. *Mamm. Genome* **4:** 303–313.

Neame, P. J., Sommarin, Y., Boynton, R. E., and Heinegård, D. (1994). The structure of a 38-kDa leucine-rich protein (chondroadherin) isolated from bovine cartilage. *J. Biol. Chem.* **269:** 1–8.

Oldberg, Å., Antonsson, P., Lindblom, K., and Heinegård, D. (1989). A collagen-binding 59-kD protein (fibromodulin) is structurally related to the small interstitial proteoglycans PG-S1 and PG-S2 (decorin). *EMBO J.* **8:** 2601–2604.

Patthy, L. (1987). Detecting homology of distantly related proteins with consensus sequences. *J. Mol. Biol.* **198:** 567–577.

Roughley, P. J., White R. J., Magny, M-C., Liu, J., Pearce, R. H., and Mort, J. S. (1993). Non-proteoglycan forms of biglycan increase with age in human articular cartilage. *Biochem. J.* **295:** 421–426.

Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S. H., Thurston, S. J., and Birkenmeier, E. H. (1994). Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm. Genome* **5:** 253–274.

Scholzen, T., Solursh, M., Suzuki, S., Reiter, R., Morgan, J. L., Buchberg, A. M., Siracusa, L. D., and Iozzo, R. V. (1994). The murine decorin: Complete cDNA cloning, genomic organization, chromosomal assignment and expression during organogenesis and tissue differentiation. *J. Biol. Chem.* **269:** 28270–28281.

Shen, Z., Heinegård, D., and Sommarin, Y. (1998). Chondroadherin expression changes in skeletal development. *Biochem. J.*, in press.

Shinomura, T., and Kimata, K. (1992). Proteoglycan-Lb, a small dermatan sulfate proteoglycan expressed in embryonic chick epiphyseal cartilage, is structurally related to osteoinductive factor. *J. Biol. Chem.* **267:** 1265–1270.

Sommarin, Y., Larsson, T., and Heinegård, D. (1989). Chondrocyte–matrix interactions. *Exp. Cell Res.* **184:** 181–192.

Vogel, K. G., Paulsson, M., and Heinegård, D. (1984). Specific inhibition of type I and type II collagen fibrillogenesis by the small proteoglycan of tendon. *Biochem. J.* **223:** 587–597.

Wahle, E. (1995). 3′ end cleavage and polyadenylation of mRNA precursors. *Biochim. Biophys. Acta* **1261:** 183–194.

Wegrowski, Y., Pillarisetti, J., Danielson, K. G., Suzuki, S., and Iozzo, R. V. (1995). The murine biglycan: Complete cDNA cloning, genomic organization, promoter function and expression. *Genomics* **30:** 8–17.

Wicki, A. N., Walz, A., Gerber-Huber, S. N., Wenger, R. H., Vornhagen, R., and Clemetson, K. J. (1989). Isolation and characterization of human blood platelet mRNA and construction of a cDNA library in lambda gt11. Confirmation of the platelet derivation by identification of GPIb coding mRNA and cloning of a GPIb coding cDNA insert. *Thromb. Haemostasis* **61:** 448–453.

Yagi, M., Edelhoff, S., Disteche, C. M., and Roth, G. J. (1995). Human platelet glycoproteins V and IX: Mapping of two leucine-rich glycoprotein genes to chromosome 3 and analysis of structures. *Biochemistry* **34:** 16132–16137.